RESEARCH ARTICLE

# Analysis of agricultural economic resilience based on big data analysis and information network technology diffusion

Xiaoyun Wu*

School of Economics and Management, Neijiang Normal University, Neijiang, Sichuan, China.

**Received:** April 3, 2024; **accepted:** May 22, 2024.

**Agriculture is an important component of the national economy, which is closely related to the country's food security and economic development. With the development of the economy, economic elasticity in agriculture gradually becomes an important research direction. This study proposed more appropriate guidance for agricultural development and national economic development based on the analysis results of agricultural economy to better understand the elasticity of it. K-means algorithm was used for complex data clustering analysis. Flink stream processing framework was introduced for effective content extraction. The differential integrated moving average model and index model were added to the data process to take into account time-series involved in agricultural economic elasticity. The information network technology was used for system resource integration and utilization. Meanwhile, interference suppression between system users was designed. The results showed that the calculation time of the proposed method only took about 180 seconds when cluster nodes were 33, while the processing time of the method was about 470 s when data size node was 400. In the identity scenario, the single thread data throughput of the method was 280,000 pieces/s. The proposed method demonstrated effective clustering and meeting normal distribution. After the method reached a convergence state, the average packet delivery rate was maintained around 99%. The relative prediction accuracy of the method was as high as 95.958%. The results indicated that the proposed method had excellent performance and could efficiently process data in the study of agricultural economic elasticity, obtaining accurate research results.**

**Keywords:** economic elasticity; big data; information network technology; K-means; autoregressive integrated moving average model.

***Corresponding author:** Xiaoyun Wu, School of Economics and Management, Neijiang Normal University, Neijiang 641100, Sichuan, China. Email: goingnba@163.com.

## Introduction

With the continuous improvement of agricultural production and economy, agricultural economic elasticity has become a research hotspot. Economic elasticity (EE) refers to the adaptability and responsiveness of an economic system to external shocks or changes [1, 2]. In agriculture, EE mainly focuses on changes in agricultural production and markets including the impact of factors such as climate change, natural disasters, and market fluctuations on agricultural output and farmers' income. Agriculture involves numerous factors and complex changes, requiring the establishment of scientific and effective data analysis and prediction models [3]. The scattered data sources, diverse data types, and uneven data quality of agricultural economy make it difficult to collect and organize data [4, 5]. Many scholars have conducted research on EE to provide a more accurate analysis of economic development. Chovancov and Tej proposed a variable decoupling approach to address the energy industry impact on EE by linking economic and environmental indicators. Meanwhile, the energy emissions impact on economic growth was evaluated. The results indicated that the proposed method effectively evaluated environmental factors in EE and provided data reference for national institutions to formulate policies [6]. Gelber *et al*. proposed a testing method for the net tax share to address the impact of taxation on EE by analyzing budget set kink change after the policy guidance knot change. The results showed that the proposed method effectively made judgments on the effective marginal tax impact in a short-term [7]. Razzaq *et al*. developed a moment quantile regression method to solve the relationship between $CO_2$ emissions and EE. By removing the normality of carbon dioxide emissions data and constructing a matrix for regression values calculation. Through practical application, the results suggested that the proposed method effectively determined the impact of carbon dioxide emissions on EE, which could be applied to the tourism industry [8]. Liang *et al*. proposed a regression framework-based prediction model for the impact of the petroleum industry on EE. Single autoregressive models were combined. Elastin net and Lasso method were applied to model search. The results found that the proposed method had good predictive performance and a certain guiding role for investment in the petroleum industry [9]. Knoblach *et al*. designed a reference effective substitution elasticity research method to address the impact of the relationship between capital and labor on EE by constructing a total substitution elastic framework of a multi-factor framework and analyzing the substitution degree of parameters. The results showed that the proposed method effectively performed EE analysis and predicted possible future trends [10]. Wang *et al*. developed an energy management model based on demand side management to address the correlation between energy terminals and EE. The model was constructed as a two-stage scheduling mode. A carbon emission flow model was used to track the virtual carbon flow. The results indicated that the proposed model effectively achieved energy detection and provided reference for the government in the energy field from the perspective of EE [11].

Statistical methods such as regression analysis and correlation analysis used in current studies of agricultural EE are more effective in dealing with small-scale data sets. However, problems such as high computational complexity and overfitting of models may be encountered in the face of large-scale and high-dimensional agricultural economic data. Econometric models are often used to assess the impact of agricultural policies or to predict market changes. However, they often require a large number of assumptions and require high data accuracy and completeness. Qualitative research methods such as case studies and expert interviews can provide in-depth insights. However, there are limitations in quantitative analysis and wide applicability. If researchers want to better study the economic resilience of agriculture, they need to design methods with better performance. Some scholars have also conducted research on big data (BD) analysis technology. Mariani *et al*. proposed an analysis model based on BD analysis technology to address the consumer discourse trends in online environments by collecting and analyzing data from two different platforms with added timeline. The results showed that the proposed method effectively analyzed the environmental discourse trends of online consumers [12]. Chen proposed an intelligent manufacturing decision-making method based on BD for resource control and decision-making during enterprise production using machine learning technology to predict production products and combinations with data collected and diagnosed during the manufacturing process. The results indicated that the proposed method accurately analyzed production data and provided visual guidance for the manufacturing process [13]. Zhang *et al*. developed an energy consumption analysis method based on BD analysis for energy consumption control in the manufacturing industry. The energy and hidden generation anomalies were detected. Data reduction and correlation analysis techniques were introduced to summarize a deep trust network-based energy consumption analysis method and found that the proposed method effectively detected energy efficiency and anomalies, providing effective data reference for energy consumption control in the manufacturing industry [14]. Zhang *et al*. proposed a reinforcement method based on BD analysis to address the generalization performance issue in image classification using a neural network with hierarchical sub-networks for iterative learning to generate global features. The results showed strong generalization performance and high classification accuracy in multiple different training samples [15]. Moreover, Gravili *et al*. proposed a BD analysis method based on a holistic framework to analyze the influencing factors of the healthcare industry. An econometric economic model was established through panel data analysis and a sample survey was conducted in more than 20 European countries. The results showed that the method effectively performed qualitative analysis and had good analytical performance in different data environments [16].

K-means is a commonly used unsupervised algorithm that can be used for agricultural data clustering to identify relationships and patterns between data. The Flink stream processing framework can achieve real-time data processing and analysis, which is of great significance for analysis and decision-making of agricultural production and market data. The differential integrated moving average model is a time series-based prediction model that can be used to predict variables such as agricultural output and market demand. The interference suppression system based on game theory can achieve system stability and optimization. This study focused on integration of BD analysis and information network technology to develop a new agricultural and EE research method to provide feasible reference solutions for agricultural EE research topics.

## Materials and methods

### Construction of agricultural EE data system based on BD analysis

The amount of data involved in agricultural EE is enormous and there are many types of data. Building an agricultural BD can provide a professional combination of agricultural markets, agricultural resource conditions, and more comprehensive data analysis [17, 18]. Before entering the database, the data was processed through three steps including extraction, conversion, and loading (Figure 1). The source of data included additional data files and other added contents in addition to the original data source. A temporary database was established to store the data during the process, reduce the amount of redundant data, and reduce system pressure. Because data in agricultural EE often have certain connections with each other, cluster analysis can better extract the connections between complex parameters [19, 20]. K-means clustering analysis has high accuracy, fast convergence speed, and good performance for large-scale data environments. The core idea of K-means algorithm is to minimize the sum of distances of each data point to the center of the cluster to which it belongs by iterative optimization. In each iteration, the algorithm first calculated the distance of each data point to the center of each cluster. Then, each data point was assigned to the cluster where the center of the cluster was closest. When all the data was moved into the cluster, an operation was completed. The Euclidean distance calculation was shown in equation 1.

$$dist(x,C) = \min_{x\in X, c\in C} \|x-c\| \quad (1)$$

where $X$ was the initial dataset. $x$ was the data in $X$. $C$ was the cluster center set. $c$ was the centralized data. The cluster center was then updated to equation 2.

$$c_i = \frac{1}{|NC_i|}\sum_{x_i\in c_i} x_i \quad (2)$$

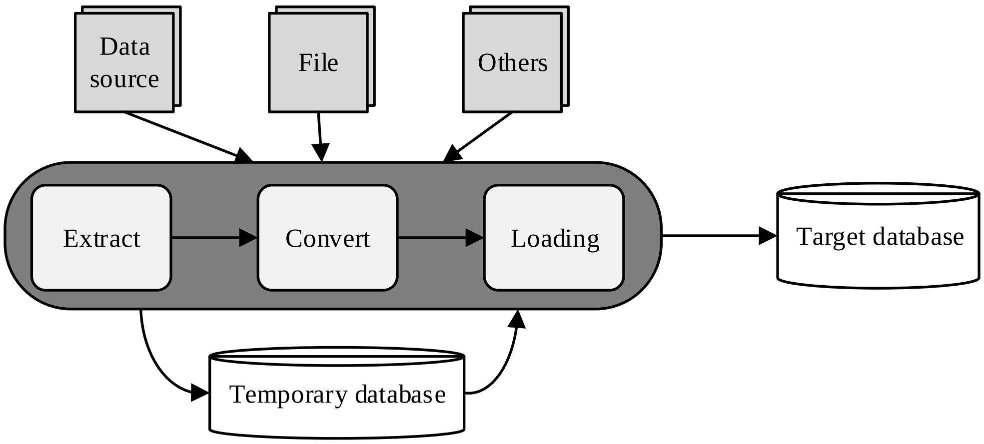


**Figure 1.** Flowchart of data processing.

where $c_i$ was the new cluster center. $NC_i$ was the data number. The sum of squared errors of the data was calculated in equation 3.

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} dist(x, C)^2 \tag{3}$$

where $E$ was the sum of square error. When this error was limited to the initial set threshold, the algorithm stopped iteration. In the agricultural BD, data heterogeneity is a serious problem, which means that data from different sources, formats, and structures cannot be directly integrated and analyzed effectively. The scheme of establishing distributed computing cluster was then designed. The effective content of discrete data could be extracted by quickly screening and filtering to improve the efficiency and accuracy of data analysis. Flink framework (https://flink.apache.org/) is an open-source stream processing framework with high throughput, low latency, and fault tolerance, which can handle large-scale real-time data flow and provide rich data processing and analysis functions and was selected as the stream processing framework [21]. When establishing a distributed computing cluster, the data must be preprocessed and cleaned to ensure the quality and consistency of the data before filtered and converted in real time using Flink's stream processing capability. Dynamic analysis and real-time monitoring of data could be realized by defining appropriate window functions and conditional expressions. The architecture of BD analysis was shown in Figure 2, where BD was mainly divided into 6 layers with the source data layer being the source of architectural data, which included log, user, operational, and other data. Different layers formed the task scheduling module of the architecture. The collection layer had two sub sections including real-time data and offline data collections. Real-time data collection monitored the log files of the database and synchronized them to the processing layer. Offline data collection achieved timed incremental data synchronization. The processing layer was also divided into two parts including real-time and offline processing. MapReduce was used for data processing. The storage layer stored the data and used a distributed file system to complete data backup. Column storage was used to segment the database, reducing the irrelevant data involved in data processing. The computing layer performed real-time and offline calculations on tasks. The access layer visualized the data processing results for easy data query and retrieval. The platform monitored hardware clusters, data computing engines, and data status. A multi-copy mode was set up for data storage to ensure data security and availability. In the study of agricultural EE, the transaction data of agricultural products market was regarded as one of the most

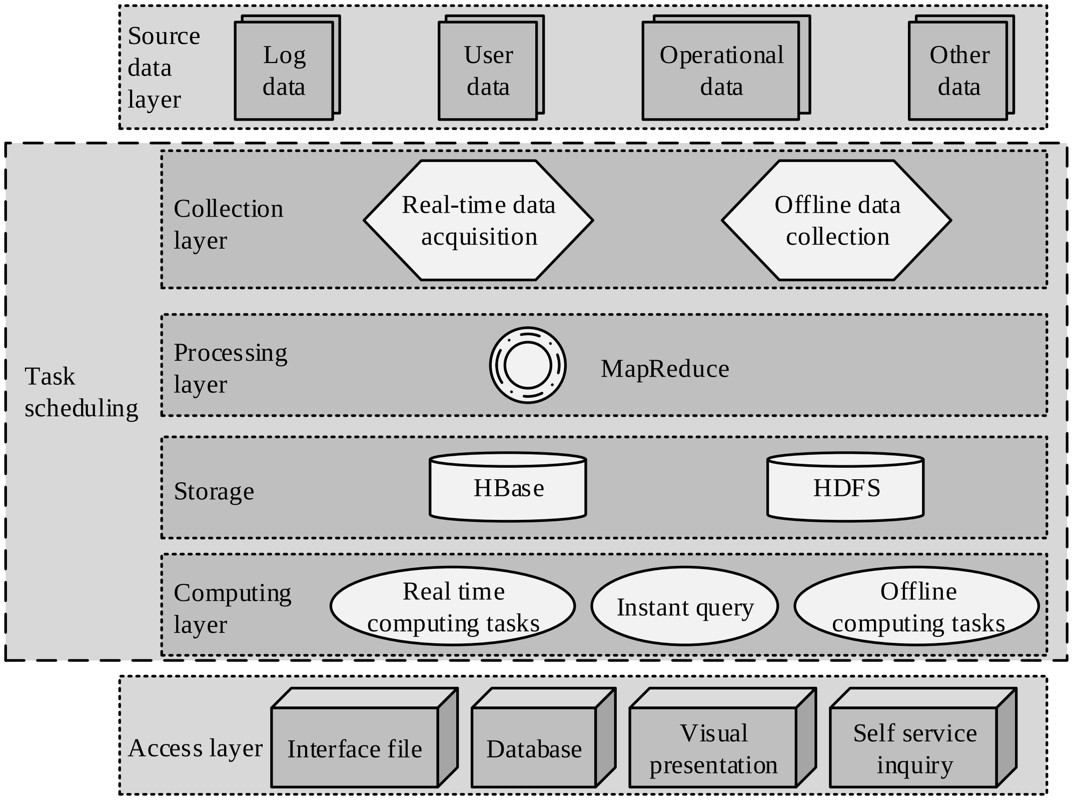


**Figure 2.** Big data analysis architecture.

important factors because the price at which agricultural product was traded was determined by a combination of factors including yield, geography, and weather. The change of these factors would directly affect the price of agricultural products, which would be in a floating state all the time [22]. The autoregressive integrated moving average (ARIMA) model was a statistical model widely used in time-series analysis, which captured long-term trends, seasonal changes, and random fluctuations in data, providing a powerful tool for data analysis. The exponential model was a time-series model based on exponential function, which could describe the growth or decline trend of data. By using the index model, the changing trend of agricultural prices could be better understood and predicted. The ARIMA model and index model were selected to analyze agricultural product data. In an autoregressive model, the sequence value could be represented by a linear combination of previous values in equation 4.

$$AR(p): Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_p Y_{t-p} + \xi_t \tag{4}$$

where $Y_t$ was the observed values in time-series. $\mu$ was the mean value. $\beta$ was the unit weight in time-series. $\xi_t$ was random perturbation in time-series. In the moving average autoregressive model, the time-series value could be represented by a linear combination of residuals in equation 5.

$$MA(q): Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q} - \xi_t \tag{5}$$

where $\varepsilon$ was the predicted residual in time-series. $\theta$ was the unit weight in time-series of the moving autoregressive model. The two combinations were mixed as shown in equation 6.

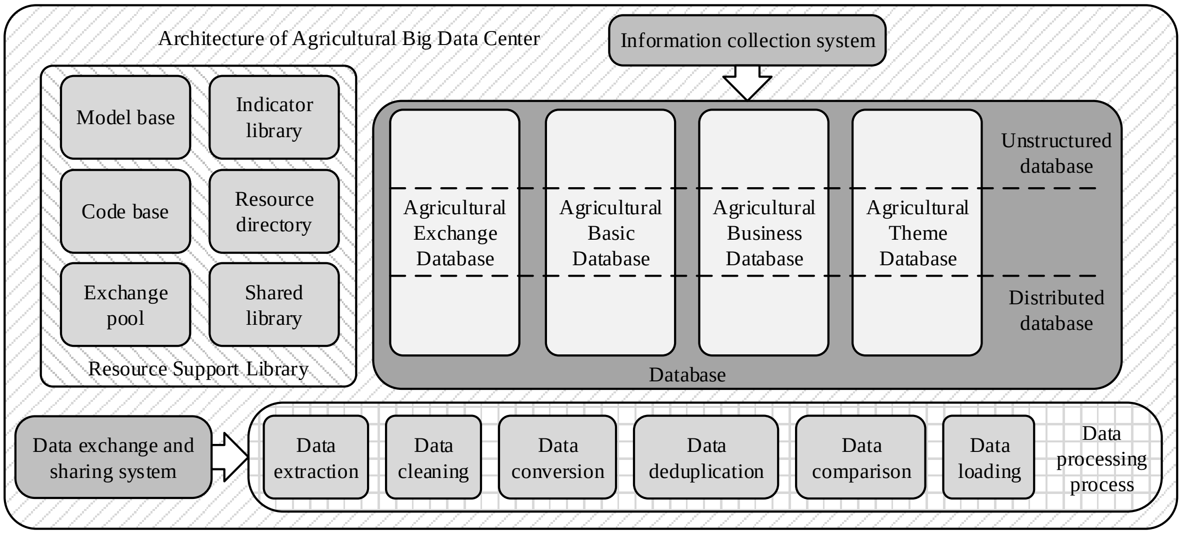


**Figure 3.** Architecture of agricultural big data center.

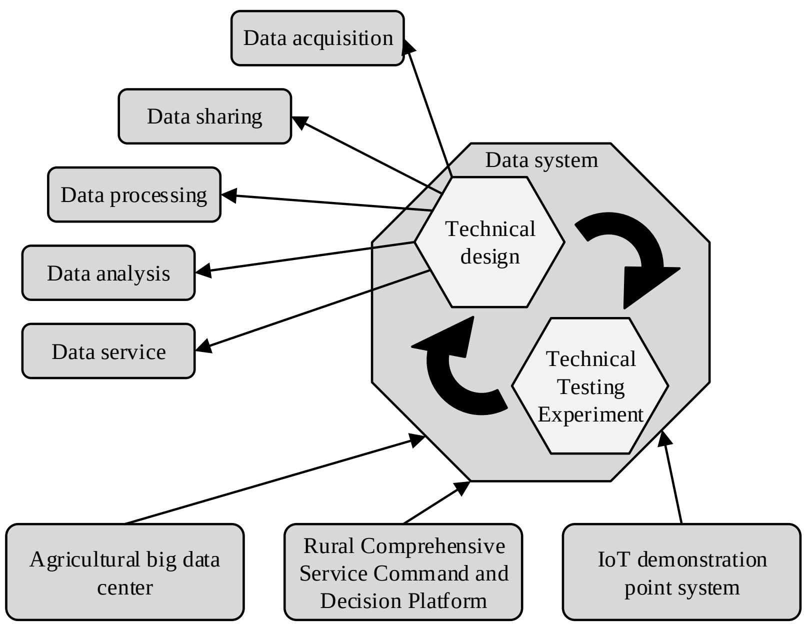


**Figure 4.** Agricultural economic elastic data system based on big data analysis.

$$AR(p): Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q} + \xi_t \quad (6)$$

where the observed values were the linear combination of past $p$ observations and $q$ residuals. Based on this, a corresponding agricultural data center architecture was constructed (Figure 3). The architecture of the agricultural BD center consisted of three parts including resource support library, data processing, and database. The information collection system served the database part, inputting the collected relevant information into the database. The database was divided into four categories based on the type of data including exchange, foundation, business, and topic. According to the characteristics of the data, they were divided into two types as unstructured and

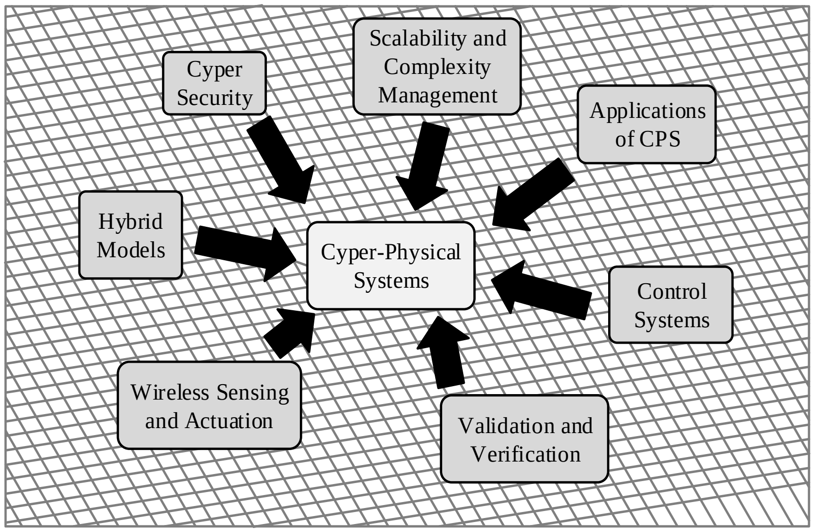


**Figure 5.** Information physics system composition.

distributed. The data exchange and sharing system served the data processing process and processes agricultural BD. A complete agricultural EE data system based on BD analysis was then constructed (Figure 4), where the agricultural EE data system was based on the agricultural BD center, rural comprehensive service command and decision-making platform, and internet of things (IoT) demonstration point system. The technical system of data included technical design and technical testing experiments. The results of the testing experiments guided the technical design to be updated. A complete data system ensured that agricultural BD could be well classified and analyzed. More accurate results could be output in subsequent EE research [23].

**Construction of agricultural EE research system based on BD analysis and information network technology diffusion**

Information network is an important technology, which has certain potential for application in agricultural EE research [24]. Information physics system is an intelligent network control system, which can integrate and innovate traditional computing methods and physical processes. Meanwhile, it can more reasonably integrate and utilize system resources. The information physics system integrated computing, communication, and control based on environmental awareness to form a controllable and scalable networked physical device system (Figure 5). Physical and computational processes interacted and provided feedback, achieving real-time interaction and deep fusion. In the study of EE, it is necessary to calculate the information content of related indicators, and only in this way, the role and value of these indicators in economic resilience can be accurately understood and assessed. The entropy method is a multi-attribute decision-making method based on information theory, which can reflect the importance of each index by calculating the information content of each index. The great advantage of this method is that it is not affected by dimensions. Therefore, the entropy method was used to calculate the amount of information [25]. If the system was in multiple different states, the system entropy was calculated using equation 7.

$$e = -\sum_{i=1}^{m} p_i \ln p_i \tag{7}$$

where $e$ was the system entropy value. $p$ was the probability of each system state. The indicator information entropy was calculated using equation 8.

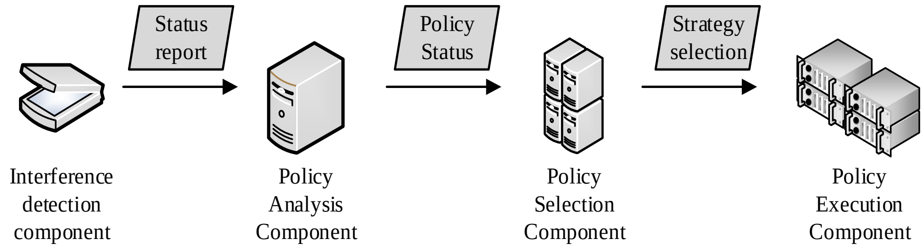


**Figure 6.** Game theory interference suppression system.

$$\begin{cases} e_j = -\sum_{i=1}^{m} p_{ij} \ln p_{ij} \\ p_{ij} = r_{ij} / \sum_{i=1}^{m} r_{ij} \end{cases} \tag{8}$$

where $e_j$ was the index information entropy when evaluation object number was $m$. $r_{ij}$ was the element in the original state probability matrix. Information entropy is a measure of the disorder of the system. The smaller the information entropy of a certain index, the greater the variation degree of the index value, the more information will be provided. Therefore, in the comprehensive evaluation, this indicator will play a larger role, which means that it should be given a higher weight. In information network technology, interference between users can lead to a large amount of data retransmission, seriously leading to packet loss. Therefore, suppressing interference between users is an important aspect [26]. An interference suppression system based on game theory was proposed and mainly composed of interference detection components, policy analysis components, policy selection components, and policy execution components (Figure 6). The detection component received the interference and generated the corresponding status report, which was then transmitted to the policy analysis component. The policy analysis component analyzed the data in status report, formulated different strategies, and transmitted them to the policy selection component. The strategy selection component judged input strategy, selected the most suitable strategy for current demand, and inputted it into strategy execution component. The policy execution component executed the received policy. In the game model, it is necessary to minimize user losses if in a non-cooperative game state. The cost function of the user was shown in equation 9.

$$u_i(\sigma_i, \sigma_{-i}) = \tau_i \left(\gamma_i^0 - \gamma_i^j\right)^2 + \xi_i p_i^j, \rho_i^j = 1 \tag{9}$$

where $\tau_i$ and $\xi_i$ were non-negative real numbers, and their ratio was used to adjust the trade-off between signal-to-noise ratio and energy consumption. $\gamma_i^0$ was the ideal signal-to-noise ratio threshold that users needed for specific situations. $\gamma_i^j$ was the signal-to-noise ratio of the user on the channel. $p_i^j$ was the transmission power of the user on the channel. The signal-to-noise ratio was calculated using equation 10.

$$\gamma_i^j = \frac{G_i p_i^j}{\sum_{k=1,k\neq i}^{M} \left( S_k p_k^j / \left(v_k^i\right)^{\delta} \right) + \eta_i} \tag{10}$$

where $v_k^i$ was the distance between two users. $\eta_i$ was the background noise power of other interference. $G_i p_i^j$ was emission intensity. The link gain was then calculated using equation 11.

$$G_i = \frac{S_i}{d_i^{\delta}} \tag{11}$$

where $d_i$ was the maximum distance between the control node and the sensing node. $S_i$ was the attenuation coefficient. When the user acted, the response method was shown below.

$$\pi_i(I_i) = \arg\min_{\sigma_i \in \wedge_i} u_i(\sigma_i, I_i), \forall b_i \in B \tag{12}$$

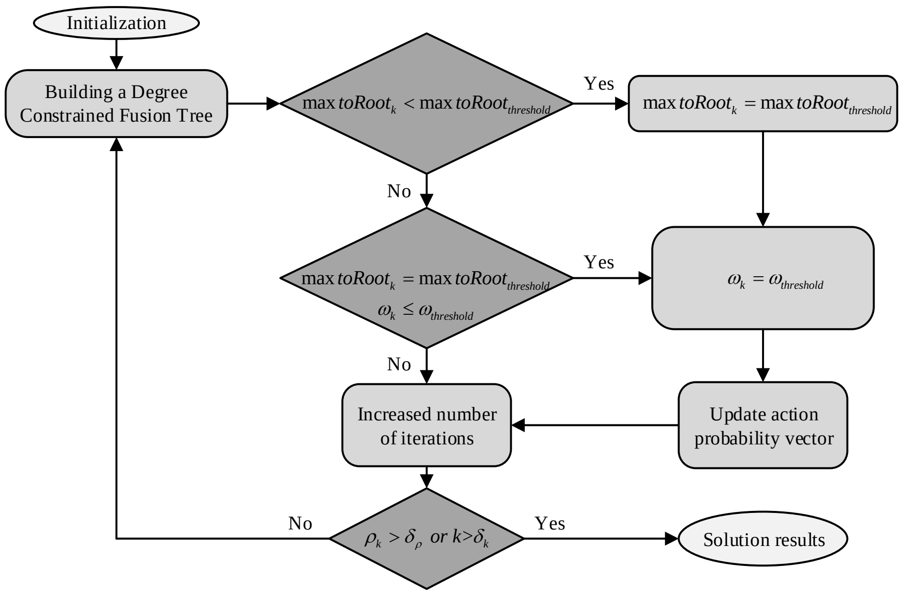


**Figure 7.** Degree constrained data fusion tree algorithm.

where $I_i$ was interference information. When transmitting data, it is necessary to ensure its reliability and real-time performance. A degree constrained real-time data fusion tree scheme should be studied and constructed. Learning automata should be integrated into the scheme. The degree constrained data fusion tree algorithm first initialized the data, and then constructed the degree constrained data fusion tree (Figure 7). Afterwards, the maximum path weight value was used to determine whether to update the weight, while completing the update of the path weight sum. The action probability vector was then recalculated. The iteration number continued to increase as progressed. The algorithm's solution result was output until the selection probability or specific threshold met the requirements. Economic development is an important component of EE. When conducting EE research, it is necessary to accurately evaluate the economic development situation and construct an evaluation index system for economic development. The evaluation indicator system follows the comprehensiveness, objectivity, and quantifiability principles to ensure the evaluation indicators scientificity and accuracy. The constructed economic development quality evaluation index system in this study contained four indicators including per capita GDP, regional GDP growth rate, per capita income, and per capita local public finance budget, which reflected the level of regional economic development, economic growth, and economic volume, respectively. Per capita GDP measured the level of economic development in a region and represented the economic output generated by each resident. The higher the per capita GDP, the higher the level of economic development of the region, and the living standard of the residents was relatively high. The GDP growth rate reflected the rate of regional economic growth. The higher growth rate meant that the economic development of the region was fast, which was conducive to raising the income of residents and improving the quality of life. Per capita income was a measure of the income earned by each resident. The higher the per capita income, the higher the living standard of the residents, and the consumption power was relatively strong. The per capita local public finance budget reflected the local government's investment in public services and infrastructure construction. A higher per capita local public finance budget indicated that local governments had invested more in improving people's well-being and promoting economic development, which was conducive to improving residents' living standards and regional development potential. The selected indicators reflected economic development, economic growth rate, and economic volume of the region, respectively, and are all objective data, ensuring that the evaluation results did not contain subjective components. To calculate the indicator weights, the original data matrix was first constructed in equation 13.

$$X = \begin{pmatrix} X_{11} & X_{12} & L & X_{1n} \\ X_{21} & X_{22} & L & X_{2n} \\ L & L & L & L \\ X_{m1} & X_{m1} & L & X_{mn} \end{pmatrix}_{m*n} \tag{13}$$

where $X$ was the original data. $m$ was the townships included in the research institute. $n$ was the evaluation indicator. Afterwards, the data range was normalized. If the indicator belonged to a positive indicator, the forward calculation method was used in equation (14).

$$X_{ij}^{*} = \left(X_{ij} - \min\left\{X_j\right\}\right) \tag{14}$$

where $X_{ij}^{*}$ was the indicator value after dimensionless initial data. $X_{ij}$ was the actual value of the indicator. $\min\left\{x_j\right\}$ was the minimum value of indicator $j$. If the indicator belonged to a negative indicator, the negative calculation method was used in equation 15.

$$X_{ij}^{*} = \left(\max\left\{X_j\right\} - X_{ij}\right) / \left(\max\left\{x_j\right\} - \min\left\{x_j\right\}\right) \tag{15}$$

where $\max\left\{x_j\right\}$ was the maximum value of indicator $j$. The adjusted proportion of the index value was then calculated. The information entropy value and the difference coefficient of the index were obtained. Meanwhile, the index weight coefficient was finally calculated. Each indicator weight coefficient could be used to calculate the corresponding evaluation value. After integrating multiple factors involved, the agricultural EE research method was formed.

**Validation of agricultural EE research system**

The performance of the constructed system was tested first to verify the validity of the research method of agricultural EE. The basic hardware environment of the experiment included 2.62 Hz CPU with Intel Core™ i7-8700K CPU @ 3.70 GHz, 16 GB PC memory, 64 GB system running memory, RTX-2080 graphics card, while the software environment included Windows 10 system (Microsoft, Redmond, WA, USA), Python (https://www.python.org), Apache HBase™ (https://hbase.apache.org), Spark framework (https://spark.apache.org), Sequel Pro (https://sequelpro.com) as database management tool, and Postman API Platform (https://www.postman.com) as testing tool. The test sample data volume was 100 Gb, which contained a data block size of 128 Mb. The sample data was obtained from various online trading platforms related to agriculture including National Agricultural Data Research Center (Beijing, China), Southern Institute of Agricultural Information Technology (Guangzhou, Guangdong, China), and Agricultural supply Chain Information Platform (Western Agricultural Science and Technology Group, Chengdu, Sichuan, China) with 1 TB, 800 GB, and 500 GB data being retrieved, respectively. Among all the data, 70% data was used for model training and the remaining 30% data was used for model validation and testing. However, the test sample data did have a small amount of duplication. When analyzing the practical application effect of the system, the data samples were the production data and trading data of agricultural products within the administrative scope of Taiyuan City, Shanxi, China. The data sources were official statistics from China Agricultural Information Network (Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing, China). The system was also compared with the Holt-Winters index calculation method that is a time-series forecasting technique combining three main components including horizontal, seasonal, and random fluctuations, enabling the analysis of sales data and economic indicators.

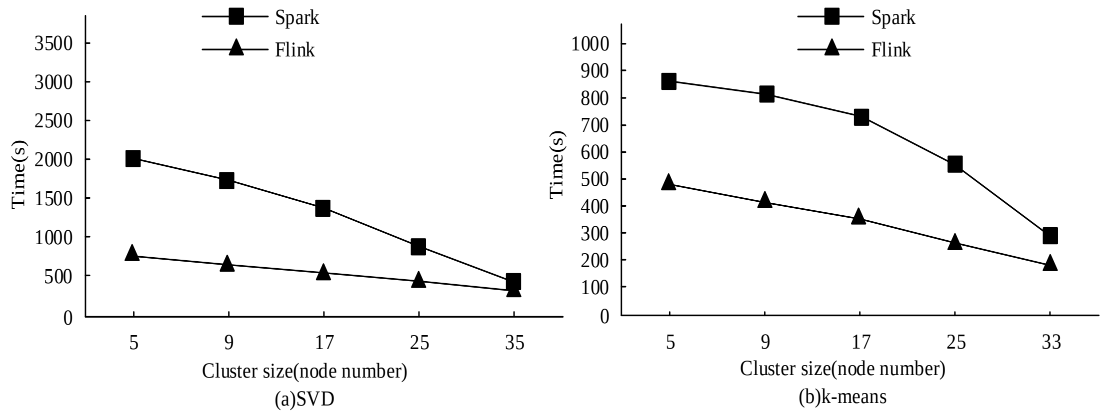


**Figure 8.** Computational performance testing.

## Results

### Computational performance testing

The replication factor for batch testing was set to 3. The calculation time was tested while data size remained unchanged and cluster size changed. When the model based on singular value decomposition and K-means used different frameworks, the computational time continuously decreased as the cluster size increased (Figure 8). When the model based on singular value decomposition calculated in clusters of different scales, the calculation time using Flink framework reached about 750 s while cluster node was 5. When cluster node reached 33, the calculation time decreased to about 340 s. When calculating in clusters of different sizes using K-means-based models, the calculation time using Flink framework reached approximately 470 s when cluster node was 5. When cluster node reached 33, the calculation time decreased to about 180 s. The calculation time of the K-means-based model using Spark framework reached approximately 860 s when the number of cluster nodes was 5. When cluster node reached 33, the calculation time decreased to about 280 s. When using Spark framework for calculation based on singular value decomposition, the calculation time reached 2,000 s when cluster node was 5. When cluster node reached 33, it decreased to about 430 s. The results indicated that the system could effectively utilize the cluster to optimize its own performance when cluster size increased and combine k-means and Flink to achieve good computational efficiency.

### Scalability testing

The calculation time of the system was tested while cluster size remained unchanged and data size changed. When the model based on singular value decomposition and K-means used different frameworks, the processing time continuously increased with the increase of data size. When a model based on singular value decomposition processed data of different scales, the time using Flink framework was about 140 s when node was 50. When the data scale node reached 400, the time increased to about 560 s. When using K-means based models for data processing with different scales, the time was about 130 s when node was 50. When node reached 400, the time increased to about 470 s. When a model based on singular value decomposition processed data of different scales, the time using Spark framework was about 200 s when node was 50. When node reached 400, the time increased to about 1,460 s. When using the Spark framework based on K-means for calculation, it took approximately 190 s when data scale node was 50. When the data scale node reached 400, the processing time increased to about 1,000 s (Figure 9). When data size increased, the processing time of the system increased more slowly, indicating better framework scalability.

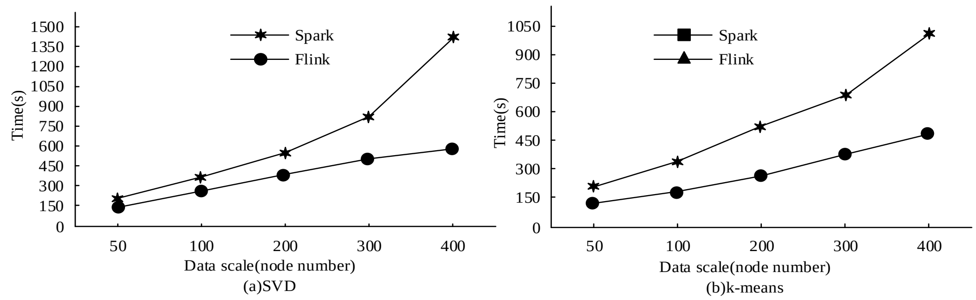


**Figure 9.** Scalability testing.

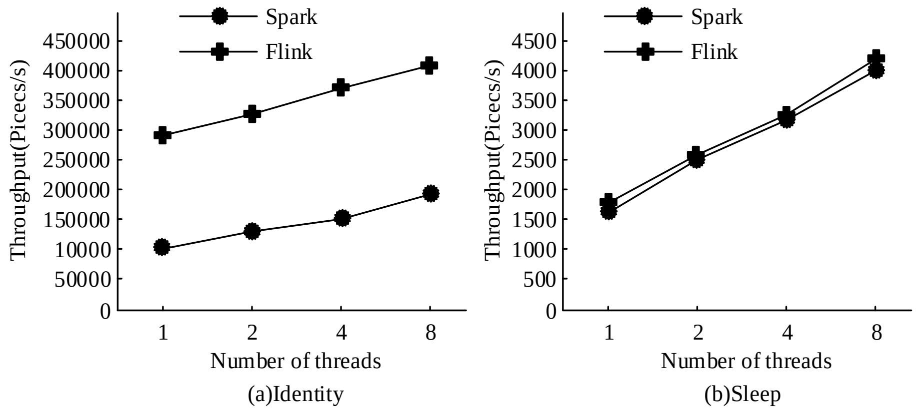


**Figure 10.** Throughput testing.

**Throughput testing**

The data throughput of the system was tested with two scenarios including identity and sleep. The identity scenario was a simple data processing and a time-consuming scenario. The results showed that data throughput of both methods increased as thread increased in different scenarios (Figure 10). In the identity scenario, the data throughput using the Spark framework was approximately 100,000 pieces/s when thread was 1. When thread reached 8, the data throughput increased to approximately 180,000 threads/s. The data throughput using the Flink framework was approximately 280,000 pieces/s when thread was 1. When thread reached 8, the data throughput increased to approximately 410,000 threads/s. In the sleep scenario, the data throughput using the Spark framework was approximately 1,700 threads/s when thread was 1. When the thread reached 8, the data throughput increased to approximately 4,000 threads/s. When using the Flink framework, the throughput was approximately 1,800 threads/s when the thread was 1. When the thread reached 8, the data throughput increased to approximately 4,200 threads/s. The data throughput of the system could effectively increase with thread count increasing, indicating the better data throughput capability of the system.

**Clustering effect testing**

The clustering effect of this system was tested, and the results were visualized in Figure 11. After visualization of the unclassified coordinates, the data in the altitude layer were loosely distributed between 3-10 intervals. The data in the flight time layer were loosely distributed between 30-180 intervals. After entering the output temperature layer, data began to aggregate towards the vicinity of 60 and -30. After entering the output pressure layer at the end, the data

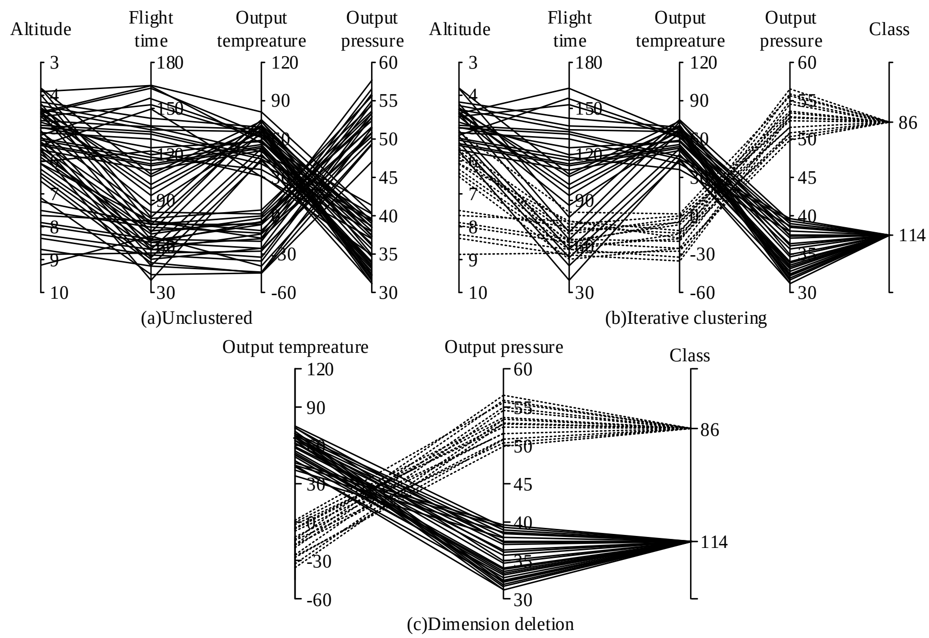


**Figure 11.** Clustering effect testing.

split towards both ends. A data vacuum zone appeared within the 42-47 range. After iterative clustering, the data were divided into two categories in the altitude layer and in a loose distribution state. After entering the flight time layer, one type of data had gradually gathered, while the other type of data was still loosely distributed. After entering the output temperature layer, the two types of data were distinguished, starting to gather around 60 and -30, respectively. After entering the output pressure layer, the two types of data were differentiated towards both ends and finally entered the class layer. The two types of data were clustered to positions 86 and 114, respectively, to complete cluster analysis. After dimension deletion, the altitude layer and flight time layer were deleted. Leaving the output temperature layer, pressure layer and class layer were output. The data in these layers were effectively clustered, which showed that the system could be effectively clustering with intuitive and simple visual results.

**Model fitting diagram**

Model fitting results showed that the model fitting data points output by the two functions of qqnorm and qqline basically fell on the line, indicating that the data met the normal distribution (Figure 12).

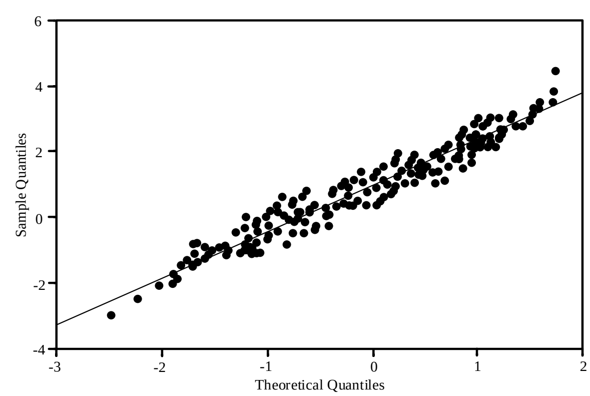


**Figure 12.** Model fitting diagram.

**Average signal-to-noise ratio, packet delivery rate, average power consumption**

The average signal-to-noise ratio was between 44 – 78 dB in the initial state. After reaching the convergence state, the average signal-to-noise ratio was maintained between 88 – 90 dB. The average packet delivery rate was between 94 - 97% in the initial state. After reaching the

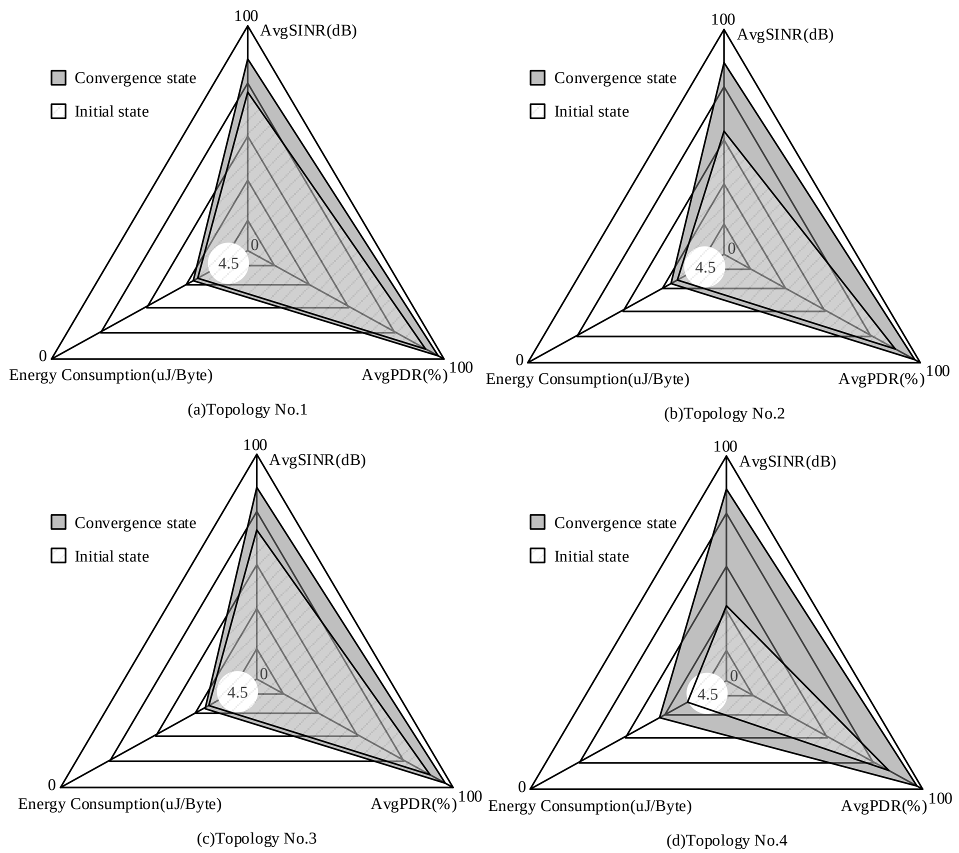


**Figure 13.** Average signal-to-noise ratio, packet delivery rate, average power consumption.

**Table 1.** Evaluation of economic development quality.

| System | Index | Unit | Indicator weight (%) | Indicator score |
|---|---|---|---|---|
| Economic development | Per capita GDP | ¥ | 29.5 | 24.000 |
| | Regional gross domestic product growth rate | % | 1.3 | 55.000 |
| | Per-capita income | ¥ | 67.3 | 43.000 |
| | Per capita local public finance budget | ¥ | 1.9 | 35.000 |
| | Total | | | 37.399 |

convergence state, the average packet delivery rate was maintained around 99%. The average power consumption of the delivery was between 3.3 - 3.8 μJ/Byte in the initial state and was reduced to between 3.1 - 3.4 μJ/Byte after reaching the convergence state (Figure 13). The results indicated that the system could effectively suppress interference after reaching a convergence state, thereby improving the delivery rate of data packets and reducing delivery power consumption.

**The application effect of EE research system in agriculture**

The practical application effect of the research system was analyzed by evaluating the quality of economic development in a rural area. The results showed that the research system generated corresponding indicator weights for different indicators, while the scores obtained from the indicators were generated based on their situations (Table 1). Data used to test the calculation error and accuracy of the system were grouped based on their origins including the

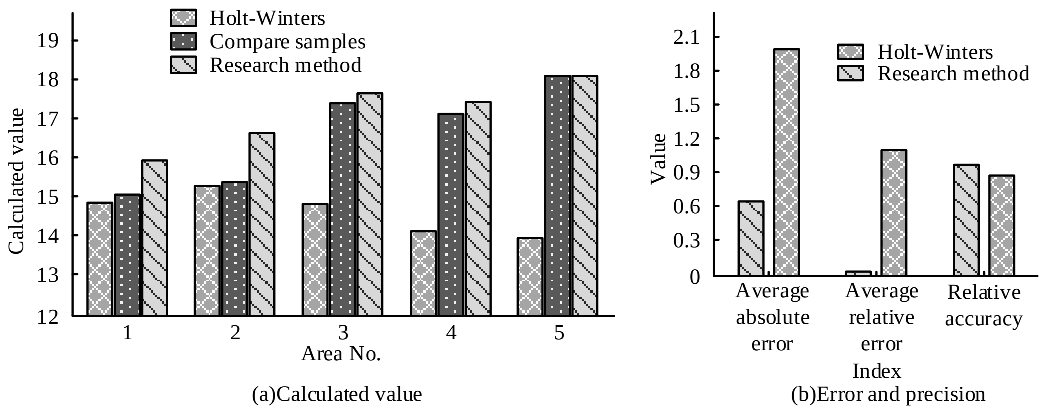


**Figure 14.** Model prediction error and accuracy.

townships of Mengfeng (1), Xugou (2), Beitian (3), Shitian (4), and Nitun (5). The results of the system were compared with the Holt-Winters index calculation method (Figure 14). The results showed that, in the regions with a comparative sample value of 15.00000, 15.33333, 17.40000, 17.10000, and 18.08333, the Holt Winters index calculation method calculated the values of 14.85374, 15.24538, 14.79171, 14.09547, and 13.93032, while the system calculated the values of 15.95204, 16.70658, 17.30468, 17.77852, and 18.15414, respectively. The average absolute and relative errors of the Holt Winters index calculation method were 2.00001 and 0.11415 with the relative precision of 88.585%. The average absolute and relative errors of the research system were 0.63401 and 0.04042 with the relative precision of 95.958%. The results indicated that the research system had better adaptability in data from different regions, and the generated calculation values had higher accuracy.

## Discussion

The study of agricultural economic resilience can help researchers predict the development trend of the agricultural economy in the future. Through the analysis of past data, possible problems can be better predicted in the future. Meanwhile, appropriate measures can be taken to deal with them. Based on the BD analysis system, this study constructed a BD system that facilitated data research on the parameters involved in agricultural economy. Information network technology was used to calculate the designed EE indicators, resulting in an agricultural EE research system based on BD analysis and information network technology diffusion. By using K-means algorithm for cluster analysis and Flink stream processing framework for content extraction, the efficiency of data processing was significantly improved. The research system used information network technology to integrate system resources and design interference suppression system. This method was helpful to optimize the operation of agricultural information system, reduce the interference and loss of data transmission, and ensure the reliability and real-time data. The results indicated that the calculation time of the research system was reduced from 470 s to 180 s when the data size remained unchanged. The computing time only increased from 130 s to 470 s with the same cluster size. The research system achieved a throughput of 410,000 threads/s in the identity scenario and 4,200 threads/s in the sleep scenario. In addition, the research system removed the altitude layer and flight time layer, successfully clustered the data into two categories, and represented these data in visual images. Two functions were applied in the system to output model fitting data points that basically fell on the line. The research system also maintained a signal-to-noise ratio between 88 -

90 dB and reduced the delivery power consumption to between 3.1 - 3.4 μJ/Byte after reaching convergence. The research system could generate index weights and calculate the evaluation results. When analyzing the application effect of five different regions, the system demonstrated the average absolute and relative errors of 0.63401 and 0.04042 with the relative accuracy of 95.958%. This high-precision forecasting ability could help agricultural producers and decision makers better understand market trends and formulate response strategies, thereby improving the adaptability and resilience of agricultural production. By introducing a differential integrated moving average autoregressive model and an exponential model, the research system was able to process and analyze time-series data, which was critical for understanding seasonal changes in agricultural production and markets as well as long-term trends. Feuerbacher *et al*. also proposed a study method of agricultural EE based on a computable general equilibrium model and analyzed the land use and the amount of agricultural chemicals [27]. The same as the research system of this study, both methods analyzed the seasonal changes of the agricultural market. However, the prediction accuracy of the method proposed by Feuerbacher *et al*. was less than 90%, which was significantly lower than that of the research system proposed in this study. The research system showed good performance in different scale data sets and cluster environments. These results indicated that the research system had good scalability and adaptability, which could cope with different scales and complexity of agricultural economic data. Liu *et al*. proposed an agricultural economic research method for resource allocation optimization, which could forecast and analyze agricultural resource pricing [28]. However, it must be combined with relevant data of water resources in the analysis and had obvious disadvantages in terms of application scope and scalability compared with the research system developed through this study. The research system of this study could be well applied to the study of agricultural EE, which could help agricultural economic management and decision-making by generating more rapid and accurate results. This research system had better data processing performance and anti-interference ability, which could generate more accurate agricultural EE research results. However, the research system was only tested in one economic policy environment. Therefore, further testing is needed in multiple economic environments to enrich experimental results and optimize the system.

## References


1. Knoblach M, Roessler M, Zwerschke P. 2020. The elasticity of substitution between capital and labor in the US economy: A meta-regression analysis. Oxford b Econ Stat. 82(1):62-82.
2. Nathaniel SP, Murshed M, Bassim M. 2021. The nexus between economic growth, energy use, international trade and ecological footprints: The role of environmental regulations in N11 countries. Energy Ecol Environ. 6(6):496-512.
3. Strickland JC, Campbell EM, Lile JA, Stoops WW. 2020. Utilizing the commodity purchase task to evaluate behavioral economic demand for illicit substances: A review and meta-analysis. Addiction. 115(3):393-406.
4. Eken S. 2020. An exploratory teaching program in big data analysis for undergraduate students. J Amb Intel Hum Comp. 11(10):4285-4304.
5. Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A. 2020. IoT, big data, and artificial intelligence in agriculture and food industry. IEEE Internet Things. 9(9):6305-6324.
6. Chovancová J, Tej J. 2020. Decoupling economic growth from greenhouse gas emissions: The case of the energy sector in V4 countries. Equilibrium. 15(2):235-251.
7. Gelber AM, Jones D, Sacks DW. 2020. Estimating adjustment frictions using nonlinear budget sets: Method and evidence from the earnings test. Am Econ J-Appl Econ. 12(1):1-31.
8. Razzaq A, Fatima T, Murshed M. 2023. Asymmetric effects of tourism development and green innovation on economic growth and carbon emissions in Top 10 GDP Countries. J Environ Plann Man. 66(3):471-500.
9. Liang C, Wei Y, Li X, Zhang X, Zhang Y. 2020. Uncertainty and crude oil market volatility: new evidence. Appl Econ. 52(27):2945-2959.
10. Knoblach M, Stöckl F. 2020. What determines the elasticity of substitution between capital and labor? A literature review. J Econ Surv. 34(4):847-875.
11. Wang Y, Qiu J, Tao Y, Zhao J. 2020. Carbon-oriented operational planning in coupled electricity and emission trading markets. IEEE T Power Syst. 35(4):3145-3157.

12. Mariani M, Borghi M. 2020. Environmental discourse in hotel online reviews: A big data analysis. J Sustain Tour. 29(5):829-848.
13. Chen M. 2022. The influence of big data analysis of intelligent manufacturing under machine learning on start-ups enterprise. Enterp Inf Syst-UK. 16(2):347-362.
14. Zhang C, Wang Z, Ding K, Chan FT, Ji W. 2020. An energy-aware cyber physical system for energy big data analysis and recessive production anomalies detection in discrete manufacturing workshops. Int J Prod Res. 58(23):7059-7077.
15. Zhang W, Wu QJ, Yang Y, Akilan T. 2020. Multimodel feature reinforcement framework using Moore–Penrose inverse for big data analysis. IEEE T Neur Net Lear. 32(11):5008-5021.
16. Gravili G, Manta F, Cristofaro CL, Reina R, Toma P. 2021. Value that matters: intellectual capital and big data to assess performance in healthcare. An empirical analysis on the European context. J Intellect Cap. 22(2):260-289.
17. Shakib M, Yumei H, Rauf A, Alam M, Murshed M, Mahmood H. 2022. Revisiting the energy-economy-environment relationships for attaining environmental sustainability: evidence from Belt and Road Initiative countries. Environ Sci Pollut R. 29(3):3808-3825.
18. Guo Y, Mustafaoglu Z, Koundal D. 2023. Spam detection using bidirectional transformers and machine learning classifier algorithms. JCCE. 2(1):5-9.
19. Shaffer B. 2020. Misunderstanding nonlinear prices: Evidence from a natural experiment on residential electricity demand. American Economic Journal: Econ Policy. 12(3):433-461.
20. Choudhuri S, Adeniye S, Sen A. 2023. Distribution alignment using complement entropy objective and adaptive consensus-based label refinement for partial domain adaptation. AIA. 1(1):43-51.
21. Zeraibi A, Balsalobre-Lorente D, Murshed M. 2021. The influences of renewable electricity generation, technological innovation, financial development, and economic growth on ecological footprints in ASEAN-5 countries. Environ Sci Pollut R. 28(37):51003-51021.
22. Nsugbe E. 2023. Toward a self-supervised architecture for semen quality prediction using environmental and lifestyle factors. AIA. 1(1):35-42.
23. Mahmud MS, Huang JZ, Salloum S, Emara TZ, Sadatdiynov K. 2020. A survey of data partitioning and sampling methods to support big data analysis. BDMA. 3(2):85-101.
24. Masood F, Masood J, Zahir H, Driss K, Mehmood N, Farooq H. 2023. Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data. JCCE. 2(1):57-67.
25. Cappa F, Oriani R, Peruffo E, McCarthy I. 2021. Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance. J Prod Innovat Manag. 38(1):49-67.
26. Talasila V, Madhubabu K, Mahadasyam MC, Atchala NJ, Kande LS. 2020. The prediction of diseases using rough set theory with recurrent neural network in big data analytics. Intell Eng Syst. 13(5):10-18
27. Feuerbacher A, Luckmann J. 2023. Labour-saving technologies in smallholder agriculture: An economy-wide model with field operations. Aust J Agr Resour Ec. 67(1):56-82.
28. Liu B, Liang Y. 2023. Impact of water resources pricing mechanism on global agricultural economy based on CGE model. Water Supply. 23(5):2135-2146.